

31759-190645  
Akihiro OKUMURA et al.  
filed 6/28/03

日 本 国 特 許 庁  
JAPAN PATENT OFFICE

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office

出 願 年 月 日

Date of Application:

2002年 6月27日

出 願 番 号

Application Number:

特願2002-187622

[ ST.10/C ]:

[ JP 2002-187622 ]

出 願 人

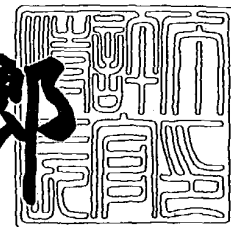
Applicant(s):

沖電気工業株式会社

2003年 4月 1日

特 許 庁 長 官  
Commissioner,  
Japan Patent Office

太田 信一郎



出証番号 出証特2003-3022440

【書類名】 特許願

【整理番号】 KN002520

【提出日】 平成14年 6月27日

【あて先】 特許庁長官 及川 耕造 殿

【国際特許分類】 G06F 15/00

【発明者】

【住所又は居所】 東京都港区虎ノ門1丁目7番12号 沖電気工業株式会  
社内

【氏名】 奥村 晃弘

【発明者】

【住所又は居所】 東京都港区虎ノ門1丁目7番12号 沖電気工業株式会  
社内

【氏名】 池野 篤司

【特許出願人】

【識別番号】 000000295

【氏名又は名称】 沖電気工業株式会社

【代表者】 篠塚 勝正

【代理人】

【識別番号】 100090620

【弁理士】

【氏名又は名称】 工藤 宣幸

【手数料の表示】

【予納台帳番号】 013664

【納付金額】 21,000円

【提出物件の目録】

【物件名】 明細書 1

【物件名】 図面 1

【物件名】 要約書 1

【包括委任状番号】 9006358

特 2002-187622

【プルーフの要否】 要

【書類名】 明細書

【発明の名称】 主要領域判定装置

【特許請求の範囲】

【請求項 1】 構造化文書から主要な領域を判定する装置において、  
定期的または不定期に構造化文書を取得する読み込み部と、  
読み込んだ構造化文書を 1 つ以上の領域に分割する分割部と、  
分割結果を一時的に記憶する分割結果記憶部と、  
領域ごとに前回の内容と今回の内容を比較して更新有無を調べる比較部と、  
領域ごとに過去の更新情報を記憶する更新情報記憶部と、  
前回までの更新情報と今回の更新有無から更新頻度を算出する更新頻度算出部  
と、

更新頻度が最も高い領域を主要領域として判定する判定部と  
を備えることを特徴する主要領域判定装置。

【請求項 2】 構造化文書から主要な領域を判定する装置において、  
定期的または不定期に構造化文書を取得する読み込み部と、  
読み込んだ構造化文書を 1 つ以上の領域に分割する分割部と、  
読み込み結果を一時的に記憶する記憶部と、  
領域ごとに前回の内容と今回の内容を比較して更新有無を調べる比較部と、  
領域ごとに過去の更新情報を記憶する更新情報記憶部と、  
前回までの更新情報と今回の更新有無から更新頻度を算出する更新頻度算出部  
と、

更新頻度が最も高い領域を主要領域として判定する判定部と  
を備えることを特徴する主要領域判定装置。

【請求項 3】 構造化文書から主要な領域を判定する装置において、  
定期的または不定期に構造化文書を取得する読み込み部と、  
読み込んだ構造化文書を 1 つ以上の領域に分割する分割部と、  
分割した領域の内容を対応するシンボルに変換する変換部と、  
各領域の内容を変換して求めたシンボルを一時的に記憶する記憶部と、

領域ごとに前回のシンボルと今回のシンボルを比較して更新有無を調べる比較部と、

領域ごとの更新情報を記憶する更新情報記憶部と、

前回までの更新情報と今回の更新有無から更新頻度を算出する更新頻度算出部と、

更新頻度が最も高い領域を主要領域として判定する判定部と

を備えることを特徴する主要領域判定装置。

【請求項4】 前記更新情報に更新頻度の値を用い、過去の更新頻度の値と今回の更新有無から決定した値との指数平均を用いて新たな更新頻度を算出することを特徴とする請求項1～3のいずれかに記載の主要領域判定装置。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】

本発明は、構造化文書の記述から主要な領域を判定し出力する主要領域判定装置に関し、例えば、WWW (World Wide Web) サイトから構造化文書を取得するシステムに適用し得るものである。

【0002】

【従来の技術】

WWWサイトに存在する構造化文書を取得し閲覧するためのツールとして、WWWブラウザがある。一般的に、構造化文書はその文書のページのレイアウト、文字の大きさなどを柔軟に指定することができるようになっている。特に、図1のように、タイトル（領域A）、他の構造化文書へのリンク（領域B）、本文（領域C）、など、ページがいくつかの領域に分割されて、WWWブラウザに表示されるような構造化文書が多く見られる。WWWブラウザを用いて、このような構造化文書から必要な情報を得るためには、ユーザは、目的の構造化文書のURLを指定し、その文書がWWWブラウザ上に表示された後に、文書をスクロールしながら目視により検索したり（人手による検索）、あるいは文字列検索機能を利用するといった作業を行なう必要がある。例えば、図1の領域Cが、ユーザの必要とする文書であったとし、こういった構造化文書が多数ある場合には、その

ユーザが必要とする情報のみを複数の構造化文書から自動的にスクラップし、1つの文書にまとめてユーザに提示されることが、人手による作業を簡略化する上で望ましくなる。このようなWWW情報抽出システムが、特開平10-187753号公報に示されている。

## 【0003】

## 【発明が解決しようとする課題】

しかしながら、上記におけるWWW情報抽出システムでは、ユーザが構造化文書中で自分が必要とするデータの開始箇所と終了箇所をあらかじめ手入力により指定することが必要である。このため、大量の構造化文書に対して実施するにはユーザの負担が大きく現実的ではなかった。

## 【0004】

本発明はこれらの問題点を解決するためになされたものであり、構造化文書における主要領域を判定することを目的とする。

## 【0005】

## 【課題を解決するための手段】

図1の例からも分かるように、主要領域以外の領域にはタイトル（領域A）や他の構造化文書へのリンク（領域B）などが存在する。これらは主要領域の記述と比較して更新頻度が低いという特徴がある。そこで、本発明では領域毎の更新頻度に基づいて主要領域を判定するように構成する。

## 【0006】

第1の本発明は、構造化文書から主要な領域を判定する主要領域判定装置において、定期的または不定期に構造化文書を取得する読み込み部と、読み込んだ構造化文書を1つ以上の領域に分割する分割部と、分割結果を一時的に記憶する分割結果記憶部と、領域ごとに前回の内容と今回の内容を比較して更新有無を調べる比較部と、領域ごとに過去の更新情報を記憶する更新情報記憶部と、前回までの更新情報と今回の更新有無から更新頻度を算出する更新頻度算出部と、更新頻度が最も高い領域を主要領域として判定する判定部とを備えることを特徴する。

## 【0007】

第2の本発明は、構造化文書から主要な領域を判定する主要領域判定装置にお

いて、定期的または不定期に構造化文書を取得する読み込み部と、読み込んだ構造化文書を1つ以上の領域に分割する分割部と、読み込み結果を一時的に記憶する記憶部と、領域ごとに前回の内容と今回の内容を比較して更新有無を調べる比較部と、領域ごとに過去の更新情報を記憶する更新情報記憶部と、前回までの更新情報と今回の更新有無から更新頻度を算出する更新頻度算出部と、更新頻度が最も高い領域を主要領域として判定する判定部とを備えることを特徴する。

## 【0008】

第3の本発明は、構造化文書から主要な領域を判定する主要領域判定装置において、定期的または不定期に構造化文書を取得する読み込み部と、読み込んだ構造化文書を1つ以上の領域に分割する分割部と、分割した領域の内容を対応するシンボルに変換する変換部と、各領域の内容を変換して求めたシンボルを一時的に記憶する記憶部と、領域ごとに前回のシンボルと今回のシンボルを比較して更新有無を調べる比較部と、領域ごとの更新情報を記憶する更新情報記憶部と、前回までの更新情報と今回の更新有無から更新頻度を算出する更新頻度算出部と、更新頻度が最も高い領域を主要領域として判定する判定部とを備えることを特徴する。

## 【0009】

## 【発明の実施の形態】

## (A) 第1の実施形態

## (A-1) 第1の実施形態の構成

図2は、第1の実施形態の主要領域判定装置の構成を示すブロック図である。第1の実施形態の主要領域判定装置は、パソコン等の通信機能を有する情報処理装置上で実現されるが、機能的には、読み込み部101、バッファ部102、分割部103、分割結果記憶部104、領域内容比較部105、更新頻度算出部106、更新頻度記憶部107及び判定部108を有する。

## 【0010】

読み込み部101は、定期的にあるいは不定期（利用者が読み込みを指示した時など）に指定された位置の構造化文書の内容をバッファ部102に読み込む機能部である。

【0011】

バッファ部102は、読み込み部101が読み込んだ内容や分割部103が処理した結果を一時的に記憶する機能部である。

【0012】

分割部103は、バッファ部102に読み込まれた構造化文書を解析することによって、予め指定された文書構造に基づいて、バッファ部102の内容を領域に分割する機能部である。

【0013】

分割結果記憶部104は、前回に分割部103が分割した結果と今回分割した結果を比較するために、前回の分割結果を記憶しておく機能部である。

【0014】

領域内容比較部105は、分割結果記憶部104が記憶している前回に読み込んだ内容の分割結果とバッファ部102が記憶している今回読み込んだ内容の分割結果とを領域ごとに内容を比較して更新の有無を検出する機能部である。

【0015】

更新頻度算出部106は、領域内容比較部105が検出した更新有無と、前回までの更新頻度から今回の更新頻度を算出する機能部である。

【0016】

更新頻度記憶部107は、分割領域ごとに内容が更新される頻度を更新頻度Sとして記憶する機能部である。

【0017】

判定部108は、更新頻度記憶部107の記憶する更新頻度に基づいて最も頻繁に内容が変化している領域を主要領域と判定し、バッファ部102から該当領域の内容を取り出してそれを出力する機能部である。

【0018】

(A-2) 第1の実施形態の動作

第1の実施形態の主要領域判定装置は、定期的にあるいは不定期（利用者が読み込みを指示した時など）に構造化文書の内容を前回と比較することにより、主要領域を判断するための情報を更新する。このため、「情報更新」時と、「主要



領域の内容出力」時の2つの動作がある。

【0019】

図3は、第1の実施形態の「情報更新」時の動作を示すフローチャートである。以下、図3のフローチャートについて説明する。

【0020】

まず、ステップS101で予め指定された位置の構造化文書の内容をバッファ部102に読み込む。このときに、他の構造化文書の挿入が必要な場合はそれを実施する。図4の構成の場合に、index.htmlを指定されたときの読み込み結果の例を図5に示す。

【0021】

次に、ステップS102でバッファ部102に読み込まれた構造化文書を予め定められた文書構造に基づいて分割し、分割したそれぞれの領域は文書の最初に表れたものから順に1から始まる領域番号を付与する。例えば、図5の内容をフレーム単位で分割する場合、処理結果（バッファ部102に書き込まれる内容）は図6のようになる。

【0022】

ステップS103では分割結果記憶部104が記憶している前回に読み込んだ構造化文書を分割した結果と、バッファ部102が記憶している今回読み込んだ構造化文書を分割部103が分割した結果とを比較して、領域ごとに更新頻度Sを算出し、更新頻度記憶部107の内容を更新する。更新頻度Sの算出方法の詳細については後述する。

【0023】

ステップS104では、バッファ部102が記憶している内容を上書きすることによって、分割結果記憶部104の内容を更新する。この内容は次の「情報更新」時に参照される。

【0024】

以上が、「情報更新」時の動作である。

【0025】

図7は、第1の実施形態の「主要領域の内容出力」時の動作を示すフローチャ

ートである。「情報更新」時のフローチャートと同じ部分は同一のステップ番号を付与し重複して説明することは省略する。

【0026】

以下、図7のフローチャートについて説明する。

【0027】

ステップS101およびステップS102は「情報更新」時のフローチャートと同じである。

【0028】

ステップS105では、更新頻度記憶部107から更新頻度Sの値が最も小さい領域番号を得る。このとき、複数の候補がある場合は領域番号の値が小さいものを優先させることにより、1つの領域番号を決定する。そして、バッファ部102から該当領域番号の内容を取り出して、これを主要領域の内容として出力し、処理を終了する。

【0029】

図8は、ステップ103で更新頻度Sを算出する方法の詳細について説明したものである。

【0030】

以下、図8のフローチャートについて説明する。

【0031】

ステップS151で、バッファ部102内の分割された領域から1つずつ領域を選択して、ステップS152からステップS159までの処理を繰り返す。

【0032】

ステップS152では、領域内容比較部105がステップS151で選択した領域の内容と、分割結果記憶部104内の同じ領域番号が示す領域の内容同士を比較する。

【0033】

ステップS153で内容を比較した結果、同じだった場合はステップS155へ、違った場合はステップS154へ進む。比較する対象が存在しない場合は、比較結果が違った場合と同様にステップS154へ進む。

【0034】

ステップS154では、バッファ部102での分割数と分割結果記憶部104での分割数を比較する。分割数が同じ場合はステップS156へ、違う場合はステップS158へ進む。

【0035】

ステップS155では、前回の内容と同じなので得点Pに100点を設定する。

【0036】

ステップS156では、前回と内容が異なっているので得点Pに0点を設定する。

【0037】

ステップS157では、指数平均によって更新頻度Sの値を求める。具体的には、更新頻度記憶部107から該当領域の更新頻度を取得し、これを前回の更新頻度S0とし、この値と得点Pの値を使って $S = S0 \cdot \alpha + P(1 - \alpha)$ を計算して求める。 $\alpha$ は $0 < \alpha < 1$ の定数で予め適当な値を決定しておく。

【0038】

ステップS158では、前回とは分割数も内容も違っているため、前回までの更新頻度との関連が不明である。このため、更新頻度Sに0を設定し初期化する。

【0039】

ステップS159では、これまでのステップで求めた更新頻度Sの値を更新頻度記憶部107の該当領域の値として更新する。

【0040】

ステップS160では、バッファ部102内の分割された領域全てに対して処理が終わるまで繰り返す。

【0041】

以上が更新頻度Sを算出する方法である。

【0042】

このような計算をすることにより、更新頻度Sは前回と内容が同じことが多い

領域ほど大きな値をとるようになり、内容が更新される頻度を表すようになる。  
つまり、値が大きいほど内容が更新される頻度が低く、値が小さいほど内容の更新が頻繁であることを示す。

## 【0043】

次に、図9および図10を使って更新頻度Sの算出の具体例を説明する。

## 【0044】

図9は分割結果記憶部104の分割数と、バッファ部102の分割数が共に3であり、領域3の内容のみ違いがある場合である。領域1, 2, 3の前の更新頻度S0をそれぞれ73, 73, 46とし、 $\alpha = 0.8$ とすると、更新頻度Sの値は以下の計算より、それぞれ78, 78, 37となる。

## 【0045】

$$73 \times 0.8 + 100 \times (1 - 0.8) \doteq 78$$

$$46 \times 0.8 + 0 \times (1 - 0.8) \doteq 37$$

図10は分割結果記憶部104の分割数が3で、バッファ部102の分割数が4であり、領域1の内容のみが同じである場合である。領域1, 2, 3の前の更新頻度S0をそれぞれ73, 73, 46とし、 $\alpha = 0.8$ とすると、内容が同じ領域1の更新頻度Sの値は上記の計算により78となる。内容が同じではない領域2～4は、更新頻度Sの値が0に初期化される。

## 【0046】

以上が第1の実施形態の動作である。

## 【0047】

「情報更新」動作を適宜実行しておくことにより、更新頻度記憶部107の内容が最適化され、「主要領域の内容出力」時に正しく主要領域を出力することができるようになる。

## 【0048】

ここまでで、「情報更新」動作と「主要領域の内容出力」動作の2種類の動作について説明したが、「主要領域の内容出力」時に同時に「情報更新」を実行しても良い。この場合、2つの動作を組み合わせ、フローチャートは図11のようになる。

【0049】

(A-3) 第1実施形態の効果

第1実施形態によれば、以下の効果を奏することができる。

【0050】

自然言語処理を用いないので、記述言語に依存せずに主要領域を判定することができる。

【0051】

構造化文書の解析を実施するが、予め指定した文書構造だけ进行处理すればよいので、全ての解析をするよりも処理が非常に軽い。

【0052】

自動的に主要領域を判定できるので、(i)指定ウェブページの更新時の通知（主要領域以外の更新は通知しない、など）、(ii)検索（主要領域以外は検索対象としない、など）、(iii)要約（主要領域のみを要約対象とする、など）などのサービスやシステムを容易に構築することが可能となる。

【0053】

分割結果記憶部104では分割した結果をそのまま記憶するので、後述する第2の実施形態のようにチェックサムを算出する必要がなく、計算量が少ない。

【0054】

(B) 第2の実施形態

(B-1) 第2の実施形態の構成

図12は、第2の実施形態の主要領域判定装置の機能的構成を示すブロック図である。

【0055】

第1の実施形態では、分割結果の内容をそのまま記憶したが、内容から算出した値（チェックサム）を記憶するように構成した。第2実施形態の構成を以下で説明する。但し、第1実施形態と同じ構成要素については、同一番号を付して繰り返し説明することは省略する。

【0056】

第2の実施形態では、第1の実施形態の分割結果記憶部104及び領域内容比

較部105に代え、チェックサム算出部201、チェックサム記憶部202及びチェックサム比較部203が設けられている。

【0057】

チェックサム算出部201は、分割結果の内容からチェックサムを算出する機能部である。

【0058】

チェックサム記憶部202は、各分割結果から算出したチェックサムの値を記憶しておく機能部である。

【0059】

チェックサム比較部203は、チェックサム記憶部202が記憶している前回のチェックサムとバッファ部102が記憶している今回のチェックサムとを領域ごとに値が同じかどうか比較する機能部である。

【0060】

(B-2) 第2の実施形態の動作

第1の実施形態と同様に、「情報更新」時と「主要領域の内容出力」時の2つの動作があるが、「主要領域の内容出力」時の動作は第1実施形態と同じなので省略し、「情報更新」時の動作についてのみ説明する。

【0061】

図13は、第2の実施形態の「情報更新」時の動作を示すフローチャートである。第1の実施形態と同じ部分は、同一ステップ番号を付与し重複して説明することは省略する。

【0062】

以下、図13のフローチャートについて説明する。

【0063】

ステップS101およびステップS102は第1の実施形態と同じである。

【0064】

ステップS201では、チェックサム記憶部202が記憶している前回の各領域のチェックサムと、バッファ部102が記憶している今回読み込んだ構造化文書の各領域の内容からチェックサム算出部201が算出したチェックサムとを比

較して、領域ごとに更新頻度 S を算出し、更新頻度記憶部 107 の内容を更新する。更新頻度 S の算出方法の詳細については後述する。

【0065】

ステップ S202 では、バッファ部 102 が記憶している分割結果からチェックサム算出部 201 がそれぞれの領域のチェックサムを算出し、チェックサム記憶部 202 の内容を更新する。

【0066】

チェックサムの算出方法としては、分割領域の内容をバイトデータ列とみなして、全てのデータを加算した結果の下位 4 バイトを取るなどとすればよい。

【0067】

この内容は次回の「情報更新」時に参照される。

【0068】

以上が、「情報更新」時の動作である。

【0069】

図 14 は、ステップ S201 で更新頻度 S を算出する方法の詳細について説明したものである。

【0070】

第 1 の実施形態と同じ部分は、同一ステップ番号を付与し重複して説明することとは省略する。

【0071】

以下、図 14 のフローチャートについて説明する。

【0072】

ステップ S251 で、バッファ部 102 内の分割された領域から 1 つずつ領域を選択して、ステップ S252 からステップ S159 までの処理を繰り返す。

【0073】

ステップ S252 では、ステップ S251 で選択した領域の内容からチェックサム算出部 201 が算出したチェックサムと、チェックサム記憶部 202 内の同じ領域番号が示す領域のチェックサム同士を比較する。

【0074】

ステップS252でチェックサムを比較した結果、同じだった場合はステップS155へ、違った場合はステップS154へ進む。

【0075】

ステップS154～ステップS160は第1実施形態と同じである。

【0076】

以上が更新頻度Sを算出する方法である。

【0077】

チェックサム同士を比較するところが第1実施形態との違いである。

【0078】

以上が第2の実施形態の動作である。

【0079】

(B-3) 第2の実施形態の効果

第2実施形態によっても、(1) 自然言語処理を用いないので、記述言語に依存せずに主要領域を判定することができる、(2) 構造化文書の解析を実施するが、予め指定した文書構造だけを処理すればよいので、全ての解析をするよりも処理が非常に軽い、(3) 自動的に主要領域を判定できるので、(i) 指定ウェブページの更新時の通知（主要領域以外の更新は通知しない、など）、(ii) 検索（主要領域以外は検索対象としない、など）、(iii) 要約（主要領域のみを要約対象とする、など）などのサービスやシステムを容易に構築することが可能となる、等の効果を奏する。

【0080】

また、第2の実施形態によれば、チェックサム記憶部202では各領域の内容から算出したチェックサムを記憶するので、第1実施形態のように、分割結果をそのまま記憶する場合と比較して、記憶容量が小さくてすむ、という効果をも奏する。

【0081】

(C) 他の実施形態

第1の実施形態で、前回との比較を実施するために分割した結果を記憶しておくように構成したが、読み込んだ内容をそのまま記憶しておき、比較する前に改



めて分割するように構成してもよい。

【0082】

第1および第2の実施形態で、分割した領域に対して文書の始めから表れた順序に従って対応付けを実施したが、構造化文書の内部でそれぞれの領域を識別することができる情報が付加されている場合は、その情報を使うように構成してもよい。

【0083】

第1および第2の実施形態で、HTMLを例に説明したが、XMLやSGMLなどでもよい。

【0084】

第2の実施形態でチェックサムを使うと説明したが、ハッシュ関数から求めたハッシュ値を使うようにしてもよい。

【0085】

1つの構造化文書进行处理する場合について説明したが、複数の構造化文書を対象としてもよい。この場合、分割結果記憶部104、更新頻度記憶部107、チェックサム記憶部202は構造化文書ごとに独立に情報を記憶する。

【0086】

第1および第2の実施形態では、構造化文書がWWWサイト上に公開されたものであることを前提としているが、記録媒体から得た構造化文書等、対象とする構造化文書の入手方法は問われないものである。

【0087】

【発明の効果】

本発明の主要領域判定装置によれば、構造化文書における主要領域を判定することができる。

【図面の簡単な説明】

【図1】

構造化文書の説明図である。

【図2】

第1の実施形態の主要領域判定装置の機能的構成を示すブロック図である。

【図 3】

第 1 の実施形態の「情報更新」時の動作を示すフローチャートである。

【図 4】

第 1 の実施形態の動作説明に用いる構造化文書の例を示す説明図である。

【図 5】

図 4 の構造化文書の所定ファイルを読み込んだ結果例を示す説明図である。

【図 6】

図 5 の構造化文書の領域の説明図である。

【図 7】

第 1 の実施形態の「主要領域の内容出力」時の動作を示すフローチャートである。

【図 8】

第 1 の実施形態の更新頻度 S の算出方法を示すフローチャートである。

【図 9】

第 1 の実施形態の更新頻度 S の算出の具体例の説明図（1）である。

【図 10】

第 1 の実施形態の更新頻度 S の算出の具体例の説明図（2）である。

【図 11】

第 1 の実施形態の「情報更新」時の動作と「主要領域の内容出力」時の動作とまとめた場合のフローチャートである。

【図 12】

第 2 の実施形態の主要領域判定装置の機能的構成を示すブロック図である。

【図 13】

第 2 の実施形態の「情報更新」時の動作を示すフローチャートである。

【図 14】

第 2 の実施形態の更新頻度 S の算出方法を示すフローチャートである。

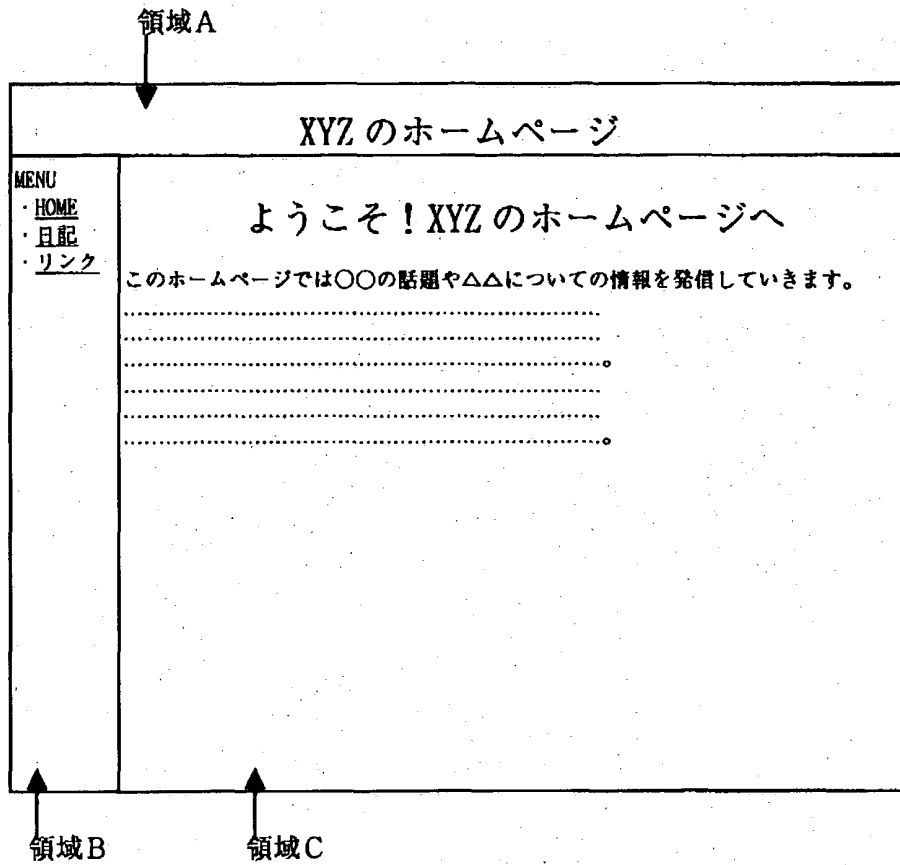
【符号の説明】

1 0 1 …読み込み部、1 0 2 …バッファ部、1 0 3 …分割部、1 0 4 …分割結果記憶部、1 0 5 …領域内容比較部、1 0 6 …更新頻度算出部、1 0 7 …更新頻

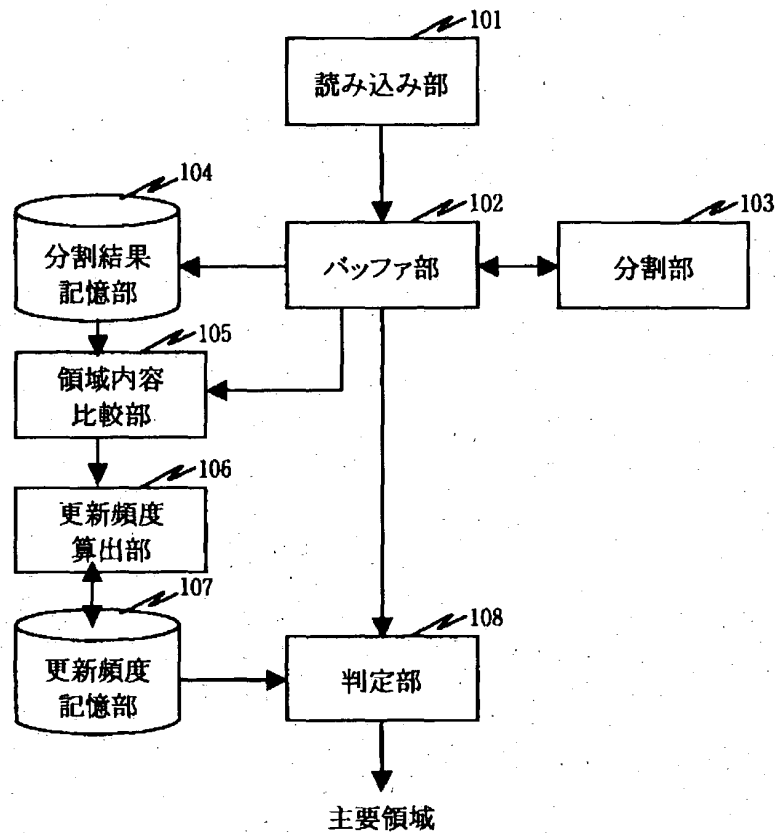
度記憶部、108…判定部、201…チェックサム算出部、202…チェックサム記憶部、203…チェックサム比較部。

【書類名】 図面

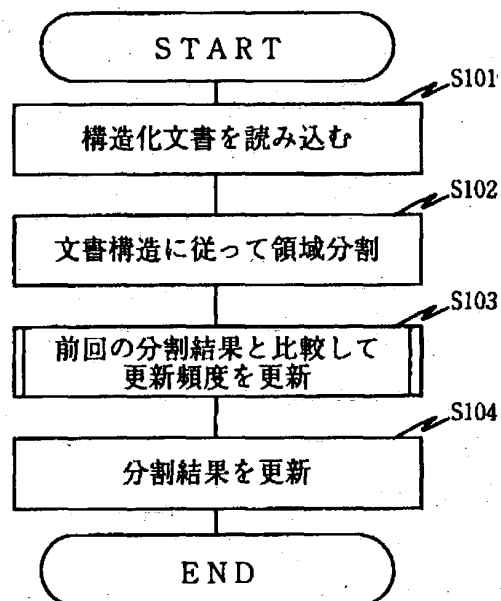
【図1】



【図 2】



【図 3】



【図4】

index.html

```

<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.0 Frames//EN">
<HTML>
<HEAD>
<TITLE>XYZ のホームページ</TITLE>
</HEAD>
<FRAMESET rows="10%,*">
<FRAME src="title.html" name="title">
<FRAMESET cols="15%,*">
<FRAME src="menu.html" name="menu">
<FRAME src="main.html" name="main">
</FRAMESET>
<NOFRAMES>
<P>フレーム非対応ブラウザをお使いの方はこちらから</P>
<P><A href="menu.htm">MENU</A>
<A href="main.htm">メイン</A></P>
</NOFRAMES>
</FRAMESET>
</HTML>

```

title.html

```

<HTML>
<HEAD>
<TITLE>タイトル</TITLE>
</HEAD>
<BASEFONT size=7 color=#008000>
<CENTER>
XYZ のホームページ
</CENTER>
</HTML>

```

main.html

```

<HTML>
<HEAD>
<TITLE>本文</TITLE>
</HEAD>
<BASEFONT size=5 color="black">
<BR>
<P>
<CENTER>
<FONT size=7>
ようこそ！XYZ のホームページへ<BR>
</CENTER>
</FONT>
</P>
このホームページでは〇〇の話題や△△についての
情報を発信していきます。
.....
.....
.....
.....
.....
</HTML>

```

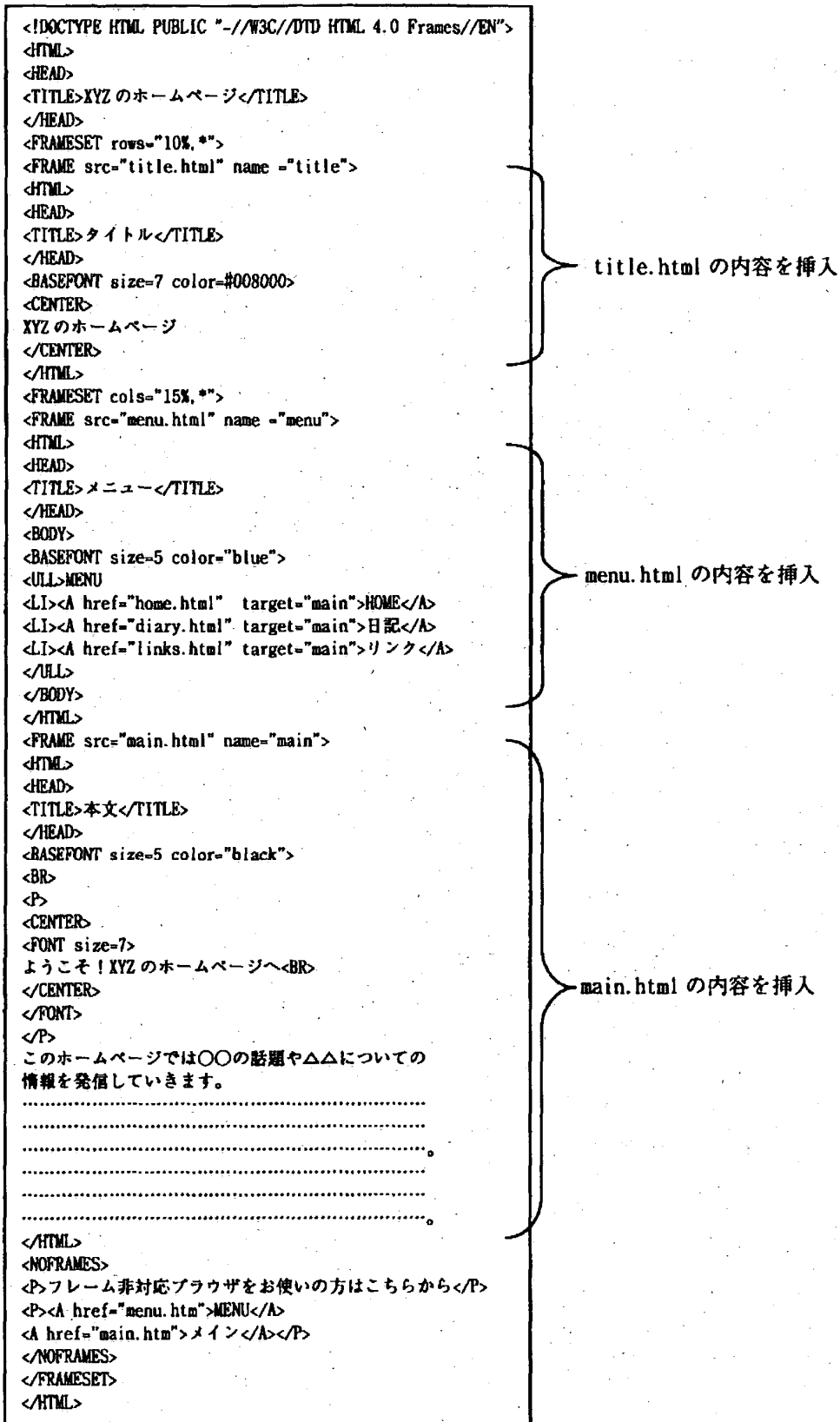
menu.html

```

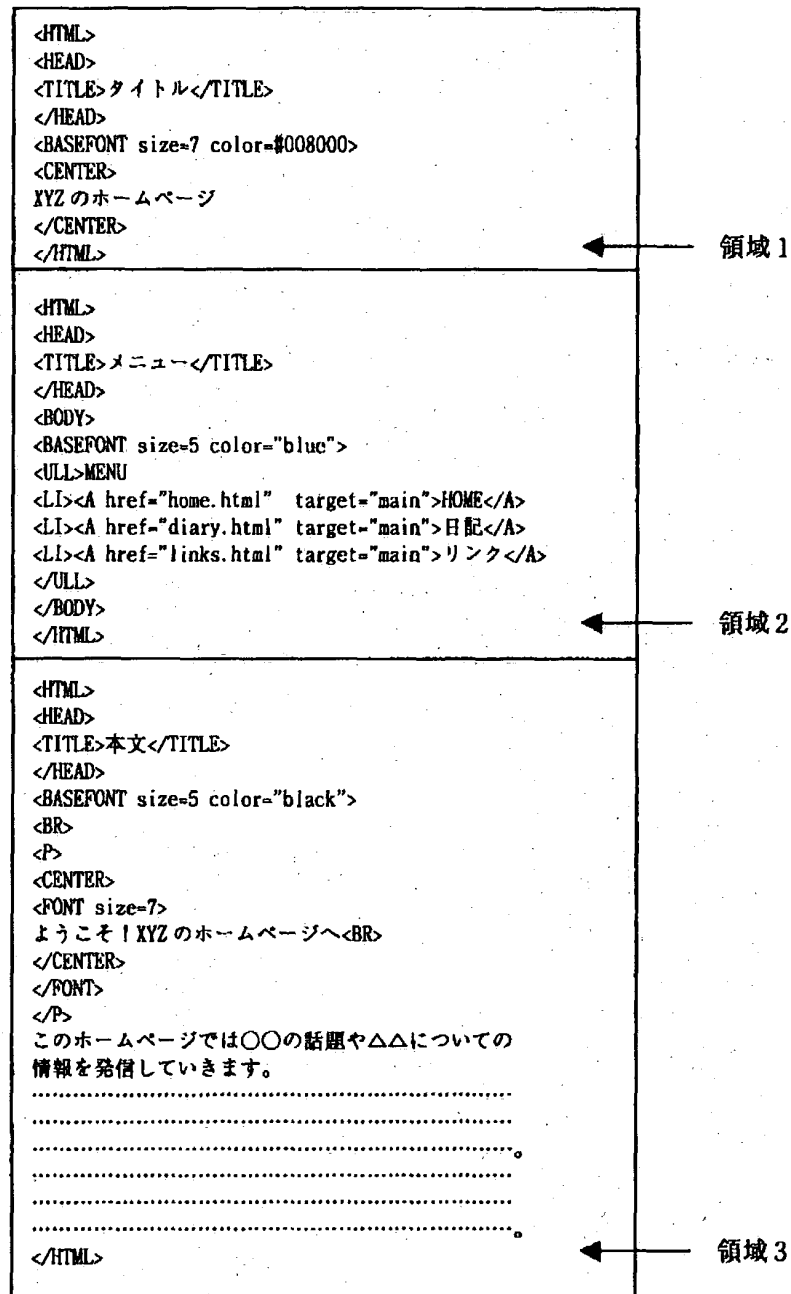
<HTML>
<HEAD>
<TITLE>メニュー</TITLE>
</HEAD>
<BODY>
<BASEFONT size=5 color="blue">
<UL>MENU
<LI><A href="home.html" target="main">HOME</A>
<LI><A href="diary.html" target="main">日記</A>
<LI><A href="links.html" target="main">リンク</A>
</UL>
</BODY>
</HTML>

```

【図 5】

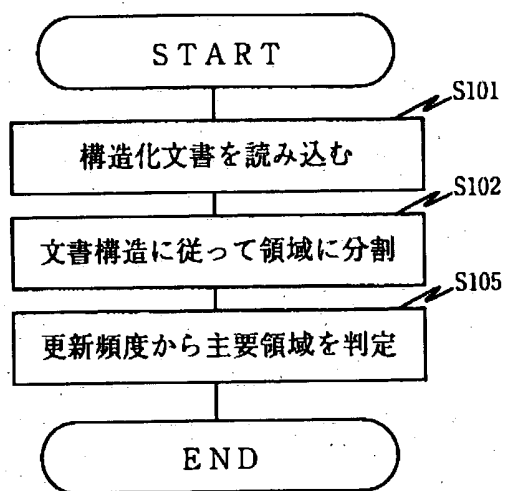


【図 6】

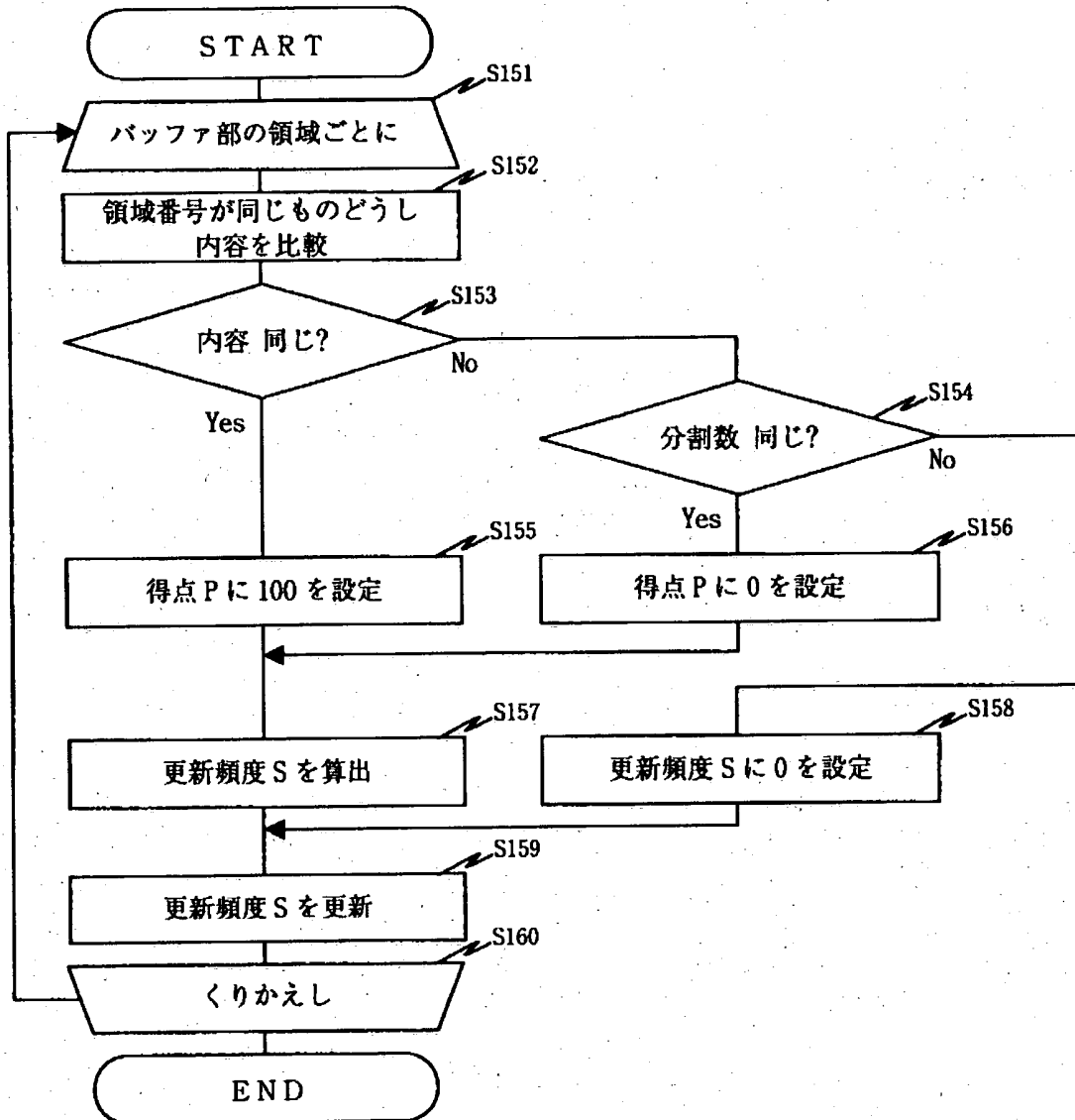




【図 7】



【図 8】



【図 9】

	比較結果	得点 P	前回の 更新頻度 $S_0$	更新頻度 S
領域 1	同じ	100	73	78
領域 2	同じ	100	73	78
領域 3	違う	0	46	37

$\alpha = 0.8$  とすると

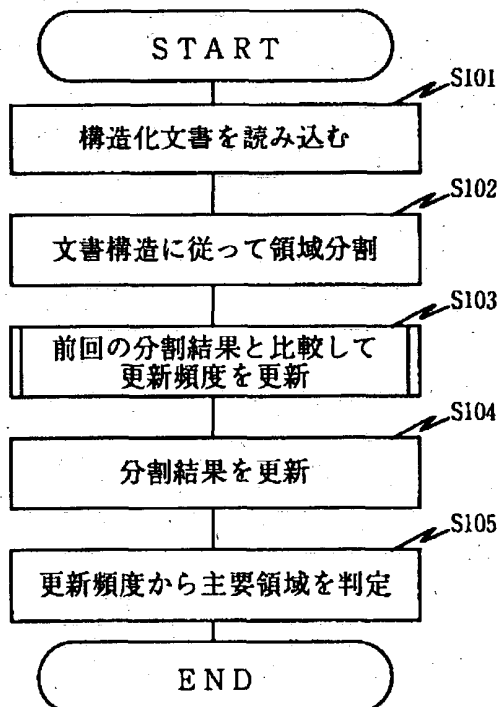
$$73 \times 0.8 + 100 \times (1 - 0.8) \doteq 78$$

$$46 \times 0.8 + 0 \times (1 - 0.8) \doteq 37$$

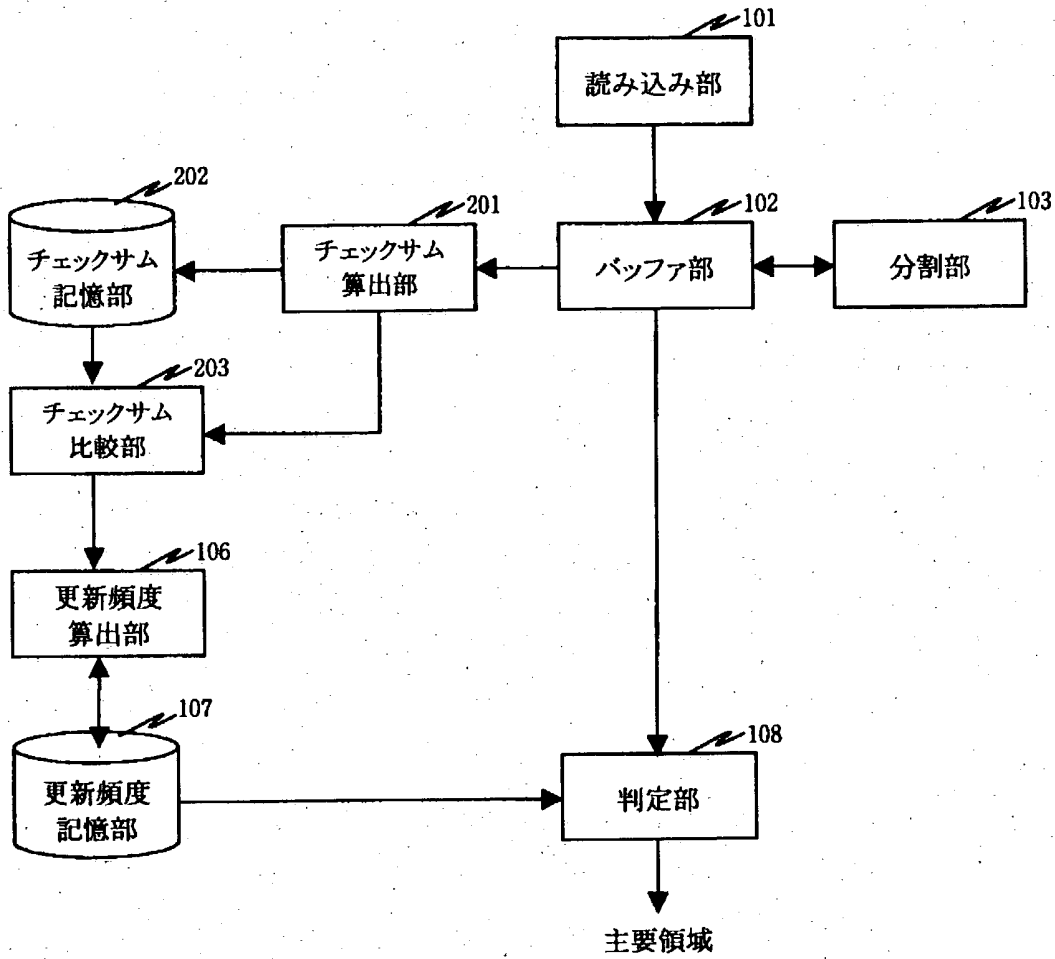
【図 10】

	比較結果	得点 P	前回の 更新頻度 $S_0$	更新頻度 S
領域 1	同じ	100	73	78
領域 2	違う	-	73	0
領域 3	違う	-	46	0
領域 4	比較対象なし	-	-	0

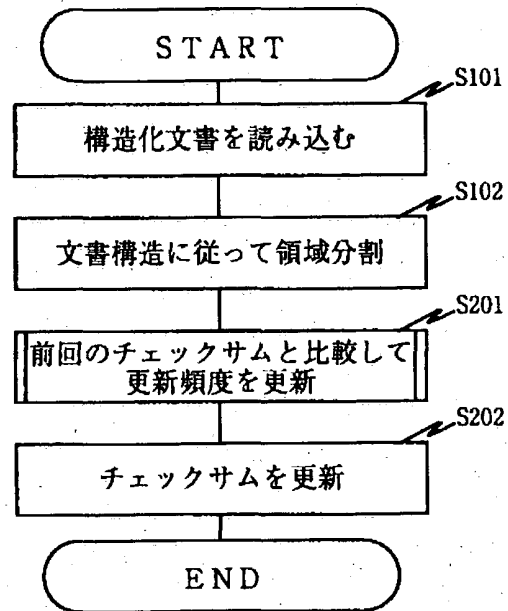
【図 11】



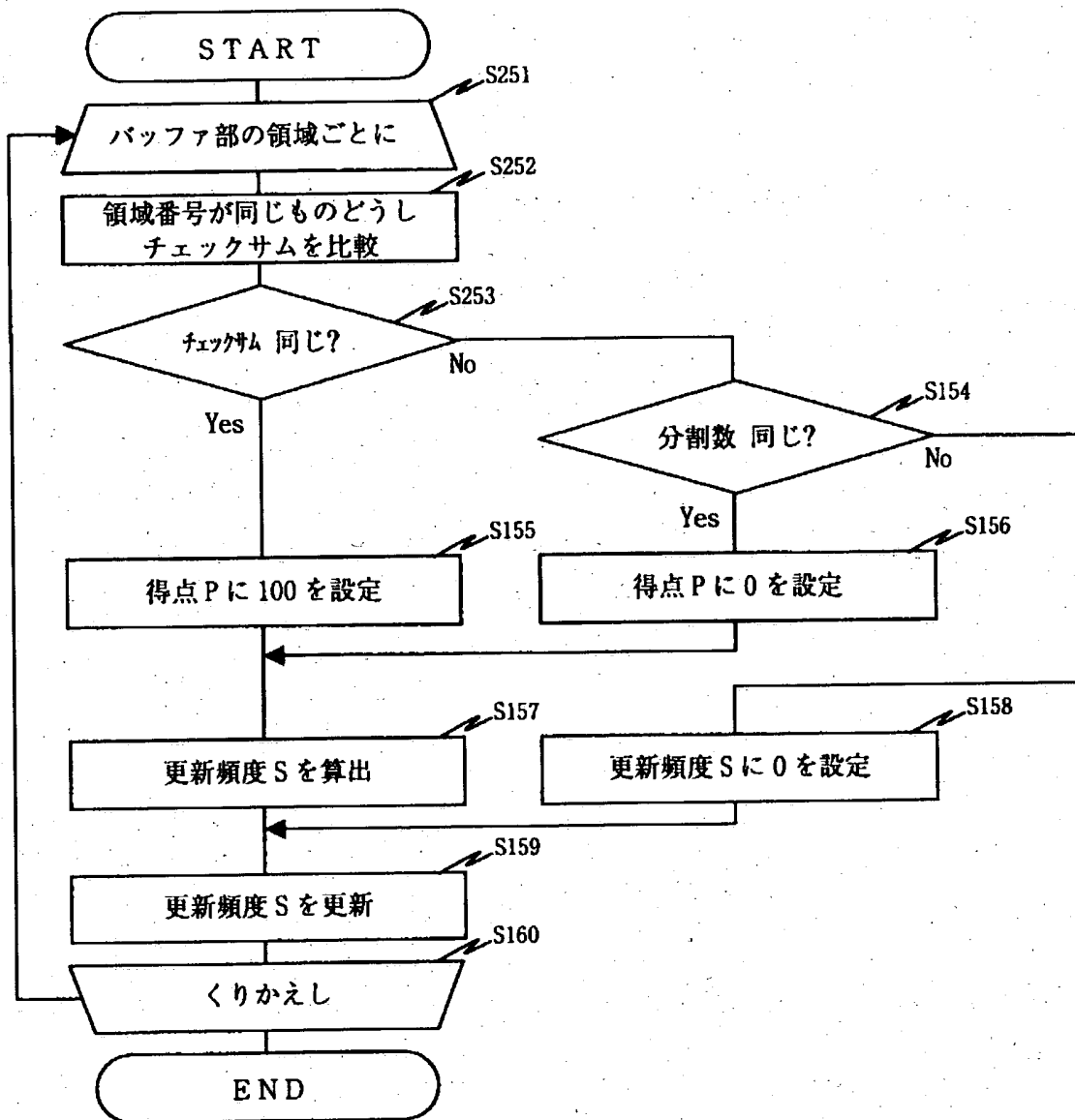
【図 12】



【図13】



【図 14】



【書類名】 要約書

【要約】

【課題】 構造化文書における主要領域を判定する

【解決手段】 本発明の主要領域判定装置は、定期的または不定期に構造化文書を取得する読み込み部と、読み込んだ構造化文書を1つ以上の領域に分割する分割部と、分割結果を一時的に記憶する分割結果記憶部と、領域ごとに前回の内容と今回の内容を比較して更新有無を調べる比較部と、領域ごとに過去の更新情報を記憶する更新情報記憶部と、前回までの更新情報と今回の更新有無から更新頻度を算出する更新頻度算出部と、更新頻度が最も高い領域を主要領域として判定する判定部とを備えることを特徴する。

【選択図】 図2

出願人履歴情報

識別番号 [000000295]

1. 変更年月日 1990年 8月22日

[変更理由] 新規登録

住 所 東京都港区虎ノ門1丁目7番12号

氏 名 沖電気工業株式会社